MATDAT18: Materials and Data Science Hackathon

Team Composition (2 people max.)

Name	Department	Institution	Email
Yihuang Xiong	Materials Science and Engineering	Penn State University	YYX5048@PSU.EDU
Weinan Chen	Materials Science and Engineering	Penn State University	WZC124@PSU.EDU

Project Title

Machine-learned model for accelerating discovery of photocatalysts

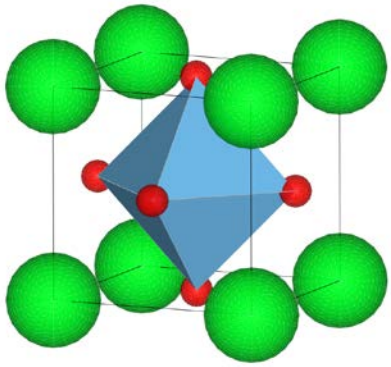
Project Synopsis (approx. 100 words)

Identified Data-Science Collaborative Need (approx. 100 words)
Data Origin and Access (*data must be available and sharable with data science teams* – please

Efficiently converting abundant solar energy into synthetic chemical fuels is of significant research interest for decades. Utilizing modern computational methods such as density functional theory enables computational screening of novel photocatalytic materials. However, predicting band edge alignment of thousands of well-equilibrated photocatalysts in aqueous environments is exceptionally demanding. Therefore, the goal of this project is to develop a statistically learned regression model that can make rapid and accurate predictions of valence band edges using physically intuitive properties of neutral elements. The developed model would not only serve as a key screening tier in the high-throughput pipeline for searching photocatalysts but will also elucidate the correlation between the atomic and electronic properties in these compounds.

address: data source/origin, access privileges, sharing privileges)

The initial structural information for photocatalytic materials belonging to the perovskite oxide family can be accessed through the database of the Materials Project.[1] The structural parameters and band edge positions in aqueous solutions are computed using open-source quantum-mechanical QUANTUM-ESPRESSO software with newly developed modules to simulate the influence of the surrounding chemical environment at low cost.[2,3] The open-source high-throughput infrastructure AiiDA will be employed to manage, record, and disseminate the database generated.[4] Data will be made available in json format as generated by AiiDA and sharable via secure online cloud services such as BOX. Materials science team members and data science team members will have privileges to manage, modify, upload and download the data.



Project Description (approx. 1.5 pages, plus figures and references; please describe data size, form, dimensionality, uncertainties, number of examples, etc.)

The production of hydrogen fuels from water splitting process has the potential to revolutionize the generation of electric power for transportation and residential applications. Hydrogen is a sustainable energy carrier whose catalytic reaction with oxygen generates electrical energy and heat without emitting carbon dioxide. Wide band-gap semiconductors with band edge positions aligned to the energy levels of water and with optical absorption compatible with the solar spectrum can enable the splitting of water molecules and the production of hydrogen fuels. One such material is $SrTiO_3$, which belongs to the materials family of cubic perovskite oxides having generic chemical formula ABO_3. These materials exhibit a wide range of attractive physicochemical properties, providing a vast combinatorial space for investigating new photocatalytic materials. Yet, only few cubic perovskite compounds have been discovered, which

fulfill the simultaneous requirements of chemical stability, band edge alignment, and optical absorbance for photocatalytic applications.

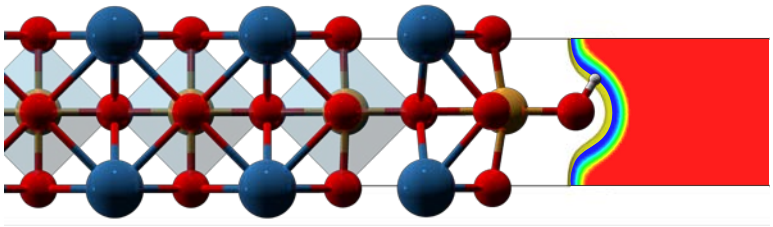


Fig 1. $SrTiO_3$ is a perovskite oxide that serves as an important reference for producing hydrogen fuels by splitting water molecules thanks to its superior photocatalytic activity. The green and red spheres represent strontium and oxygen, respectively. Titanium sits in the center of the octahedron.

The development of high-throughput combinatorial explorations based on density-functional theory (DFT) has enabled materials scientists to predict relevant photocatalytic features over extensive chemical spaces. Progress have been made in predicting new battery materials[5], energy storage polymer nanocomposites[6], solar fuels photoanode materials[7], to name a few. In spite of the predictive power of DFT, the computational screening of certain materials properties at the high-throughput level is exceptionally demanding. In the case of the rational design of photocatalytic materials, predicting valence band edges in aqueous environments is a critical tier of the high-throughput pipeline but often requires several thousands of optimized atomic geometries with strict convergence thresholds for accurate energies and interatomic forces.

Therefore, the goal of this project is to develop a statistically learned regression model that can make rapid and accurate predictions of band edges (in particular, the lowest valence band edges). Sample band edges are extracted from 500 randomly selected structures that are well equilibrated using self-consistent continuum solvation (SCCS) methods.[3] The sample size can be expanded with further DFT calculations. A range of atomic properties of neutral elements are selected as descriptors, which include but are not limited to electronegativity, ionization potential, highest and lowest occupied atomic level, valence orbital radii[8], and bond order[9]. The learned model will serve as a preselector that sieves desirable candidates from more than 2,000 cubic perovskite structures. Finally, the verified band edges will be computed on the surviving candidates to validate the accuracy of the regression and deliver the final candidate compound.

Fig 2. Symmetric construction of $SrTiO_3$ slabs using the implicit self-consistent continuum solvation method under the framework of first-principle calculations. The color gradient region corresponds to the simplified representation of the surrounding chemical environment.

The newly proposed model will deliver a robust selection scheme to explore the entire chemical space of perovskite structures. This model will not only serve as a crucial tier in the high-throughput pipeline for searching perovskite photocatalysts, but it will also deepen our chemical insights of the complex relations between the atomic and electronic properties in these compounds.

Reference

1. A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, "The Materials Project: A materials genome approach to accelerating materials innovation", APL Materials, **1**(1), 011002 (2013). DOI:10.1063/1.4812323

2. P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. Fabris, G. Fratesi, S. de Gironcoli, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, R. M. Wentzcovitch, "QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials", Journal of Physics: Condensed Matter, **21**, 395502 (2009). DOI: 10.1088/0953-8984/21/39/395502.

3. O. Andreussi, I. Dabo and N. Marzari, "Revised self-consistent continuum solvation in electronic structure calculations", Journal of Chemical Physics, **136**, 064102 (2012). DOI: 10.1063/1.3676407.

4. G. Pizzi, A. Cepellotti, R. Sabatini, N. Marzari, and B. Kozinsky, "AiiDA: automated interactive infrastructure and database for computational science", Computational Materials Science, **111**, 218-230 (2016). DOI: 10.1016/j.commatsci.2015.09.013.

5. Hautier, G., Jain, A., Ong, S.P., Kang, B., Moore, C., Doe, R. and Ceder, G., "Phosphates as lithium-ion battery cathodes: an evaluation based on high-throughput ab initio calculations." Chemistry of Materials, **23**(15), 3495-3508 (2011). DOI: 10.1021/cm200949v.

6. Shen, Z.H., Wang, J.J., Lin, Y., Nan, C.W., Chen, L.Q. and Shen, Y., "High-Throughput Phase-Field Design of High-Energy-Density Polymer Nanocomposites." Advanced Materials, **30**, 170438 (2017). DOI: 10.1002/adma.201704380 .

7. Yan, Q., Yu, J., Suram, S.K., Zhou, L., Shinde, A., Newhouse, P.F., Chen, W., Li, G., Persson, K.A., Gregoire, J.M. and Neaton, J.B., "Solar fuels photoanode materials discovery by integrating high-throughput theory and experiment." Proceedings of the National Academy of Sciences, **114**(12), 3040-3043 (2017). DOI: 10.1073/pnas.1619940114 .

8. Pilania, G., Mannodi-Kanakkithodi, A., Uberuaga, B.P., Ramprasad, R., Gubernatis, J.E. and Lookman, T., "Machine learning bandgaps of double perovskites." Scientific Reports, **6**, 19375 (2016). DOI:10.1038/srep19375

9. J. Jon Paul , Kulik. H. J., "Predicting electronic structure properties of transition metal complexes with neural networks", Chemical Science, **8**, 5137-5152 (2017). DOI:10.1039/C7SC01247K.